

**PROTEOME-WIDE MAPPING OF POST-TRANSLATIONAL
MODIFICATIONS USING ENDOPEPTIDASES
ENDONUCLEASES**

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Patent Application No.
5 60/405,589, filed August 14, 2002, the disclosure of which is incorporated herein in its
entirety for all purposes.

**STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY
SPONSORED RESEARCH AND DEVELOPMENT**

[0002] The present invention was supported by a grant from the National Institutes of
10 Health (CA ⁷⁰³³¹ 70031). The Government may have rights in this invention.

BACKGROUND OF THE INVENTION

[0003] Protein post-translational modification is one of the dominant mechanisms of
information transfer in cells. A major goal of current proteomic efforts is to generate a
system level map describing all the sites of protein post-translational modification. Recent
15 effort toward this goal has focused on developing new technologies for enriching and
quantitating phosphopeptides. By contrast, identification of the sites of phosphorylation
typically relies exclusively on the use of tandem mass spectrometry to sequence individual
peptides.

[0004] Much of the complexity of higher organisms is believed to reside in the specific
20 post-translational modification of proteins (Venter *et al.*, *Science*, 2001, 291(5507): 1304-
51.). Protein phosphorylation is the most ubiquitous such modification; almost 2% of the
human genome encodes protein kinases and an estimated one-third of all proteins contain a
covalently bound phosphate group (Manning *et al.*, *Science*, 2002, 298(5600): 1912-34).
Due to the importance of protein phosphorylation in regulating cellular signaling events,
25 there is intense interest in developing technologies for mapping phosphorylation events on a
proteome-wide scale.

[0005] Existing approaches for phosphorylation site mapping rely almost exclusively on
the use of tandem mass spectrometry (MS/MS) to sequence individual peptides in order to
localize sites of phosphorylation. Despite the power of this approach, MS/MS of
30 phosphopeptides remains challenging due to (i) the signal suppression of phosphate

BRIEF SUMMARY OF THE INVENTION

[0008] The present invention provides novel ^{endopeptidases} ~~endonucleases~~ for use in mapping post-translational modification sites in a genome, such as the human genome. The present invention provides ^{endopeptidases} ~~endonucleases~~ that, surprisingly, site-specifically cleave a post-translationally modified polypeptide at a site of post-translational modification.

[0009] In a first aspect, the invention provides a method of mapping the sites of polypeptide post-translational modifications. The method includes site-specifically cleaving a peptide bond of the post-translationally modified polypeptide with an endopeptidase at a site of post-translational modification to produce a degraded post-translationally modified polypeptide. After cleavage at the site of post-translational modification, the site of post-translational modification is determined.

[0010] In another aspect, the present invention provides an endopeptidase that site-specifically cleaves a peptide bond of a post-translationally modified polypeptide at a site of post-translational modification, wherein the endopeptidase comprises an active site that binds to said post-translational modification.

[0011] In another aspect, the endopeptidases of the present invention are produced by a method that includes introducing one or more point mutations into a model endopeptidase at one or more candidate amino acid positions in an active site of the model endopeptidase to produce a plurality of candidate endopeptidases. At least one of the plurality of the candidate endopeptidases is an endopeptidase of the present invention that site-specifically cleaves a peptide bond of a post-translationally modified polypeptide at a site of post-translational modification. The endopeptidase that site-specifically cleaves at said site of post-translational modification is identified by contacting each of the plurality of candidate endopeptidases with the post-translationally modified polypeptide to determine whether or not each candidate endopeptidase site-specifically cleaves the peptide bond of the polypeptide at the site of a post-translational modification.

[0012] In another aspect, the present invention provides an isolated nucleic acid encoding a endopeptidase which site-specifically cleaves a peptide bond of a post-translationally modified polypeptide at a site of post-translational modification and which comprises one or more point mutations at one or more amino acid positions within the endopeptidase active site. The isolated nucleic acid contains a subsequence having at least 70% nucleic acid sequence identity to a nucleic acid sequence of Figure 2.

[0013] In another aspect, the present invention provides an isolated nucleic acid encoding a endopeptidase which site-specifically cleaves a peptide bond of a post-translationally modified polypeptide at a site of post-translational modification and which comprises one or more point mutations at one or more amino acid positions within the endopeptidase active site. The isolated nucleic acid hybridizes under highly stringent hybridization conditions to a nucleic acid sequence of Figure 2, wherein the hybridization reaction is incubated at 42°C in a solution comprising 50% formamide, 5x SSC and 1% SDS, and washed at 65°C in a solution comprising 0.2x SSC and 0.1% SDS.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0014] Figure 1 is an amino acid sequence of a subtilisin model endopeptidase^(SEQ ID No:1)_Λ
- [0015] Figure 2 is a nucleic acid sequence that encodes a subtilisin model endopeptidase^(SEQ ID No:2)_Λ
- [0016] Figure 3 illustrates a comparison of a computer generated three-dimensional structure of the model subtilisin and a phosphotyrosine polypeptide.
- [0017] Figure 4 illustrates the phosphotyrosine site-specificity of candidate subtilisin endopeptidases and the model subtilisin endopeptidase against either an unmodified tyrosine or phenylalanine.
- [0018] Figure 5 shows kinetic data for the site-specific cleavage at a phosphotyrosine by a subtilisin endopeptidase containing the substitution point mutations P129G and E156R. ^(SEQ ID No:3)_Λ ^{Abz-Phe-Arg-Pro-Xxx-Gly-Phe-Y(No2)-Asp =}
- [0019] Figure 6 shows kinetic data for the site-specific cleavage at a phosphotyrosine by a subtilisin endopeptidase containing the substitution point mutations G127S and E156R.
- [0020] Figure 7 is an amino acid sequence of a subtilisin model endopeptidase containing a signal sequence (in bold) and a pro-domain (underlined)^(SEQ ID No:4)_Λ
- [0021] Figure 8 is a nucleic acid sequence that encodes a subtilisin model endopeptidase containing a signal sequence (in bold) and a pro-domain (underlined)^(SEQ ID No:5)_Λ

DETAILED DESCRIPTION OF THE INVENTION

- [0022] In contrast to presently utilized methods of developing a system level map describing all the sites of post-translational peptide modification, e.g., polypeptide phosphorylation, the present invention provides an approach for post-translational modification mapping that makes it possible to enzymatically interrogate a protein sequence directly to identify sites of post-translational modification.

[0045] "Polypeptide" refers to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a "peptide." The terms "peptide" and "polypeptide" encompass proteins. Unnatural amino acids, for example, β -alanine, phenylglycine and homoarginine are also included under this definition. Amino acids that are not gene-encoded may also be used in the present invention. Furthermore, amino acids that have been modified to include reactive groups may also be used in the invention. All of the amino acids used in the present invention may be either the D - or L - isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see*, Spatola, A. F., in CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983).

[0046] A "degraded post-translationally modified polypeptide" refers to the polypeptide fragments produced by site-specifically cleaving a post-translationally modified polypeptide at a site of post-translational modification using an ^{endopeptidase} ~~endonuclease~~ of the present invention.

[0047] The term "fragmentation pattern" refers to the configuration of the polypeptide fragments of the degraded post-translationally modified polypeptide as visualized or produced by an analytical method. A variety of analytical methods may be used to provide a fragmentation pattern. For example, where the analytical method is mass spectrometry, the fragmentation pattern is referred to as a "mass spectral fragmentation pattern." Where the analytical method is two-dimensional electrophoresis, the fragmentation pattern is referred to as a "two-dimensional electrophoretic fragmentation pattern."

[0048] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical

video display and a keyboard, a modem, an ISDN terminal adapter, an Ethernet port, a punched card reader, a magnetic strip reader, or other suitable I/O device.

[0142] The invention also preferably provides the use of a computer system, such as that described above, which comprises: (1) a computer; (2) a stored bit pattern encoding a collection of peptide sequence specificity records obtained by the methods of the invention, which may be stored in the computer; (3) a comparison post translationally modified polypeptide target; and (4) a program for alignment and comparison, typically with rank-ordering of comparison results on the basis of computed similarity values.

Kits

[0143] The present invention also provides a kit for practicing a method set forth herein. In an exemplary embodiment, the kit includes one or more component useful to practice the method of the invention and instructions for using that component to practice the method of the invention.

[0144] In a preferred embodiment, the kit includes a container of an endopeptidase for the present invention and instructions for using the endopeptidase to determine sites of post-translationally modification on the polypeptide. The examples that follow are intended to further illustrate the invention not to limit the scope of the invention.

[0145] The terms and expressions which have been employed herein are used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding equivalents of the features shown and described, or portions thereof, it being recognized that various modifications are possible within the scope of the invention claimed. Moreover, any one or more features of any embodiment of the invention may be combined with any one or more other features of any other embodiment of the invention, without departing from the scope of the invention. For example, the ^{endopeptidases} ~~endonucleases~~ described in the ^{endopeptidase} ~~endonuclease~~ section are equally applicable to the informatics methods described herein. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

EXAMPLES

Materials

[0146] The BG2036 protease deficient strain of *B. subtilis* and the pSS5 shuttle vector containing the subtilisin BPN' gene were employed. All pNA tetrapeptide substrates were

[0155] The resulting test polypeptides are shown in Figure 5, wherein Xxx represents a phosphotyrosine, sulfonyl tyrosine, tyrosine, phenylalanine, phosphoserine, phosphothreonine, alanine, valine, leucine, isoleucine, aspartic acid, glutamic acid, arginine, or lysine as shown. The data in panel A was obtained using a test polypeptide containing a succinyl-paranitroanilide fluorogenic donor-acceptor pair. The data in panel B was obtained using a test polypeptide containing an aminobenzoic acid-tyrosine(NO₂)-aspartic acid fluorogenic donor-acceptor pair.

Example 4

[0156] Example 4 demonstrates a method for identifying an endopeptidase that site-specifically cleaves a peptide bond of a post-translationally modified polypeptide. The methods involve assaying the candidate subtilisins of Example 2 with the test polypeptides of Example 3.

[0157] Kinetics for the fluorogenic substrates of the series Abz-Phe-Arg-Pro-Xxx-Gly-Phe-Y(NO₂)-Asp_A^(SEQ ID No: 3) were measured in 50 mM Bicine, 2 mM CaCl₂, pH 8.5 at 25° C by monitoring fluorescence at 420 nm upon excitation at 320 nm using a instrument. Initial rate data from 8 substrate concentrations bracketing the K_M was measured in triplicate and fit directly to the Michaelis Menten equation using the Prism software package (GraphPad,). When it was not possible to saturate the enzyme, values for k_{cat}/K_M were obtained from initial rates at low concentrations (10[S]<K_M) using the relationship k_{cat}/K_M = V_o[S]. Kinetics for tetrapeptide substrates of the series Suc-Ala-Ala-Pro-Xxx-pNa were measured by monitoring the change in absorbance at 412 nm over time using a Uvikon spectrophotometer. Protein concentrations were determined spectrophotometrically using an extinction coefficient of 32.2 mM⁻¹ cm⁻¹ at 280nm (Matsubara, 1965).

Example 5

[0158] Example 5 demonstrates that subtilisin endopeptidases that site-specifically cleave a phosphotyrosine polypeptide at the phosphorylated tyrosine are obtained using the methods of the present invention, as demonstrated in Examples 1-4.

[0159] Figure 4 illustrates the phosphotyrosine site-specificity of the candidate subtilisin endopeptidases and the model subtilisin endopeptidase against either an unmodified tyrosine or phenylalanine. As shown in Figure 4, subtilisin endopeptidases containing the following substitution point mutations were found to preferably cleave at the phosphotyrosine residue over a tyrosine residue or phenylalanine residue: G127S and E156R, P129G and E156R,

**PROTEOME-WIDE MAPPING OF POST-TRANSLATIONAL
MODIFICATIONS USING ^{ENDOPEPTIDASES} ~~ENDONUCLEASES~~**

ABSTRACT OF THE DISCLOSURE

The present invention provides novel ^{endopeptidases} ~~endonucleases~~ that site-specifically cleave a post-translationally modified polypeptide at a site of post-translational modification. The present invention further provides methods making and using the ^{endopeptidases} ~~endonucleases~~, including methods of mapping post-translational modifications in the human genome.

60006600 v1